Homotopy Methods in Infinite Dimensional Optimization

Michael Gertz

Argonne National Laboratory

April 5, 2002

M. Weiser, P. Deuflhard. The Central Path towards the Numerical Solution of Optimal Control Problems. ftp://ftp.zib.de/pub/zib-publications/reports/ZR-01-12.ps.Z.

Notation

We wish to minimize $J(x)$ subject to $c(x) = 0$ and $g(x) \geq 0$.

The functional $J : X \mapsto \Re$ where X is a (real Banach) space of functions.

The function $c : X \mapsto Z_c$ and $g : X \mapsto Z_g$ where Y and Z are also spaces of real-valued functions.

The relation $g(x) \geq 0$ is understood to hold componentwise and almost everywhere. For a discretization, this means it holds at the grid points.

Optimality

Let the Lagrangian be

$$L(x, \lambda, \eta) = J(x) - < \lambda, c(x) > - < \eta, g(x) >$$

The optimality conditions are

$$\begin{aligned} \nabla_x L(x, \lambda, \eta) &= 0 \\ c(x) &= 0 \\ < \eta, g(x) > &= 0 \\ \eta,\ g(x) &\geq 0 \end{aligned}$$

The relation $\eta \geq 0$ is understood in terms of normal cones.

In a reasonable discretization $\eta \geq 0$ is enforced at the grid points.

If X is finite dimensional, then complementarity holds pointwise. In other words $\eta_i g_i(x) = 0$.

One may solve the nonlinear system of equations

$$\begin{aligned} \nabla_x L(x, \lambda, \eta) &= 0 \\ c(x) &= 0 \\ \eta_i g_i(x) &= 0, \text{ for each } i \end{aligned}$$

while enforcing the bounds to attempt to find a minimizer.

It is difficult to manage constraints of the form $\eta_i g_i(x) = 0$, so these are typically replaced by $\eta_i g_i(x) = \mu$ with $\mu > 0$.

The path of solutions $(x(\mu), \lambda(\mu), \eta(\mu))$, when it exists, is called the *central path*.

Smoothed NCP Functions

Rather than solving $\eta_i g_i(x) = 0$ and enforcing $\eta_i, g_i(x) \geq 0$, find a zero of a function that enforces both sets of conditions.

Consider the Fischer-Burmeister function

$$\psi(a, b) = a + b - \sqrt{a^2 + b^2}.$$

If $\psi(a, b) = 0$ then $ab = 0$ and $a, b \geq 0$, but $\psi(a, b)$ is non-differentiable at the solution.

We may use a parameter $\mu > 0$ to create a smooth function

$$\psi(a, b; \mu) = a + b - \sqrt{a^2 + b^2 + 2\mu}.$$

Why we care about smoothed NCP functions

- Smoothed NCP functions are defined at every point.

- Methods that enforce positivity (interior-point) methods, can be sensitive to the starting point.

- Start strategies for interior point methods can push the starting point away from the initial value.

- Methods based on smoothed NCP functions may become competitive with interior-point methods.

The Big Picture

Let $F(x, \lambda, \eta; \mu) = (\nabla_x L(x, \lambda, \eta); c(x); \Psi(g(x), \eta; \mu))$.

Choose $\mu_0 > 0$.

Find an approximate solution $F(x, \lambda, \eta; \mu_k) = 0$.

Choose a positive $\mu_{k+1} < \mu_k$.

Use existing information to guess a solution to $F(x, \lambda, \eta; \mu_{k+1}) = 0$.

Repeat until converged.

Summary of Results

Under suitable conditions:

- The central path does exist in function space and may be followed to a solution.
- An inexact Newton iteration may be used to find a zero of $F(x, \lambda, \eta; \mu_k) = 0$
- A inexact predictor exists that may be used to guess an answer to $F(x, \lambda, \eta; \mu_{k+1}) = 0$.

The Bad News

The theory, as developed, puts fairly strong restrictions on the class of problems that may be solved.

The example in the technical report is for a problem outside this class. The authors acknowledge that the theory doesn't apply to this problem.

Details needed to implement a practical algorithm have not yet been published.

Dual Variables Revisited

Consider the Lagrangian

$$L(x, \lambda, \eta) = J(x) - < \lambda, c(x) > - < \eta, g(x) >$$

and recall $c : X \mapsto Z_c$ and $g : X \mapsto Z_g$.

Optimality only requires that

$$< \lambda, \cdot >: Z_c \mapsto \Re \text{ and } < \eta, \cdot >: Z_g \mapsto \Re$$

be bounded linear functionals. In other words $\lambda \in Z_c^\star$ and $\eta \in Z_g^\star$.

Further consideration of structure is needed to determine that multiplier exist (a.e.) as functions.

State and Control Variables

Typically, the variables x partition into two sets, the state variables y and control variables u.

For a typical control problem, the state variables are exactly those that appear on the left hand side of an ODE, $\dot{y} = f(y, u, t)$.

A state constraint is a constraint of the form $h(y, u, t) \geq 0$.

While most dual variables may be represented as functions, it is typical for dual variables associated with state constraints to be measure-valued.

The theory in this paper does not apply to state constraints.

Inexact Predictors

Let $\tau = -\log \mu$ and recall that $v = (x, \lambda, \eta)$. Let $v(\mu)$ be the central path.

The Euler predictor is $v_{k+1} = v(\text{exp}(-\tau_k)) + \Delta\tau \frac{dv}{d\tau}(exp(-\tau_k))$.

The change of variables from μ to τ is meant to encourage μ to be reduced by a factor, rather than a difference.

The authors work out the theory of this predictor in the presence of inexact evaluation of the functions.

The authors work out a means of estimating the maximum allowable size of $|\Delta\tau|$.

Inexact Newton Methods

Because Newton's method is begin done in function space, every iteration involves a discretization error.

Let $v = (x, \lambda, \eta)$ The inexact Newton method is

$$\begin{aligned} F'(v_k)\Delta v_k &= -F(v_k) + r_k \\ v_{k+1} &= v_k + \Delta v_k, \end{aligned}$$

where r_k is an (unknown) residual vector.

The authors develop the theory of inexact Newton methods in function space.

They further develop estimators for the maximum allowable size of r_k.

Adaptive Mesh Refinement

To obtain sufficient accuracy in the Newton iteration (sufficiently small r_k), the authors use adaptive mesh refinement.

The scheme used in M. Weiser's thesis was to estimate discretization error by comparing with a higher order polynomial discretization.

In numerical examples, adaptive refinement concentrates grid points around discontinuities in the solution and other "interesting" phenomena.

Affine Invariant Norms

An affine invarient norm is a family of norms $\{\|\cdot\|_v\}$ where the norm may depend on the current point v.

These norms are required to have appropriate invariance properties under certain classes of linear transformations of the variables and/or the range space.

Much of the theory in this paper in developed in terms of affine invariant norms.

While invariance may be useful, the benefit of using these norms was not demonstrated in the paper, so I won't discuss these norms further.